Genetic techniques used in the diagnosis of inherited platelet disorders

Andrew D. Mumford, MB ChB PhD FRCPath. a.mumford@bristol.ac.uk

Sarah K. Westbury, BM BCh PhD. sarah.westbury@bristol.ac.uk*

School of Cellular and Molecular Medicine

University of Bristol, UK

* corresponding author.

Research Floor Level 7, Bristol Royal Infirmary, Upper Maudlin Street, Bristol, BS2 8HW.

Tel. +44 117 342 2336 or +44 7733 011527

Word count – abstract: 98

Word count – main text: 4211

Key words:

Platelets

Genetics

Diagnosis

High-throughput DNA sequencing

Short title: Genetic diagnosis of inherited platelet disorders

ABSTRACT

Recent advances in genetic analysis are bringing huge benefits to patients with rare genetic disorders, including those with inherited disorders of platelet number and function. Modern clinical hematological practice now has a range of genetic techniques available to enable the precision diagnosis of inherited platelet disorders. There are some features of this disparate group of inherited disorders that present specific challenges to establishing an accurate genetic diagnosis. This review aims to introduce the techniques that are relevant for the genetic diagnosis of inherited platelet disorders and will discuss the key considerations necessary for their application to the clinic.

INTRODUCTION

Recent advances in the molecular diagnostic laboratory are bringing huge benefits to patients with rare genetic disorders, including inherited platelet disorders (IPD; disorders of platelet number and/or function). Hematology clinicians now have a range of genetic techniques available for diagnosis of IPD in individuals and pedigrees. However, there are some features of IPD that present challenges to genetic diagnosis. This review aims to introduce to the techniques that are relevant for genetic diagnosis of IPD and highlights some important considerations necessary for clinical application.

The relevance of achieving a genetic diagnosis in IPD

IPD are large group of genetically heterogeneous disorders, which may have similar clinical and laboratory characteristics to common acquired platelet disorders. Moreover, the different IPD are often difficult to distinguish from each other, yet may vary in heritability and prognosis, and may require different interventions.

One immediate benefit from diagnosis at genetic level is that it enables more precise classification of IPD, enabling better prediction of clinical risk of bleeding and syndromic comorbidities. For example, in pedigrees with autosomal dominant thrombocytopenia, genetic diagnosis of *CYCS*-related thrombocytopenia (*CYCS*) usually indicates favourable outcomes because there are no known associated features other than thrombocytopenia and low risk of bleeding[1]. In contrast, diagnosis of familial platelet disorder with a predisposition to acute myeloid leukaemia (*RUNX1*), which may be clinically

indistinguishable from *CYCS*-related thrombocytopenia at presentation, has different implications because bleeding is more likely as platelets are dysfunctional as well as reduced in number. Moreover, there is an approximately 40% risk of developing myeloid malignancy[2]. Detecting some types of IPD may inform selection of specific treatments, such as the use of TPO receptor agonists to temporarily correct thrombocytopenia in *MYH9*-related disorder (*MYH9*-RD)[3,4] and *DIAPH1*-related disorder[5]. Moreover, correctly distinguishing heritable thrombocytopenia from immune thrombocytopenia reduces the likelihood of ineffective and potentially dangerous therapies such as corticosteroids, intravenous immunoglobulins and splenectomy. In addition to clinical benefits to index cases with IPD, genetic diagnosis also benefits pedigree members by allowing robust detection of disease when phenotype testing is difficult and by enabling prediction of risk of disease in subsequent generations.

Phenotypic challenges and genetic heterogeneity in IPD

The widely-adopted phenotype-driven diagnostic pathway for IPD typically comprises clinical evaluation, then initial laboratory tests such as the full blood count, examination of the peripheral blood smear or bone marrow, platelet function testing and/or measurement of platelet surface protein expression by flow cytometry. This may be followed by more specialised investigations, such as examination of platelet ultrastructure by electron microscopy or specific functional tests such as measurement of dense or alpha granule release[6], which may only be available in reference laboratories or research centres.

For a minority of IPD in which there are distinctive clinical or laboratory characteristics, systematic evaluation of phenotype may sometimes be sufficient for diagnosis of a specific disorder to the level of abnormal protein or protein complex. These IPD generally are associated with severe, highly penetrant phenotypes such as Glanzmann thrombasthenia (*ITGA2B*, *ITGB3*), or unmistakable syndromic presentations such as Chediak-Higashi syndrome (*LYST*) in which there are characteristic laboratory features. However, for most patients with IPD that lack a distinctive presentation, this approach is insufficient for precision diagnosis and at best, enables detection of defects at cellular or biological pathway level only[7,8].

Efforts to standardise precision diagnosis in IPD are further hampered by the marked genetic heterogeneity of the known IPD. Considering the 56 currently known genes (as at June 2018) associated with IPD (Table 1), variants in a number of different genes can be associated with similar phenotypes observed in the clinic or phenotyping laboratory. For example, the association of autosomal dominant macrothrombocytopenia and sensorineural deafness previously thought to be pathognomonic of *MYH9*-RD is now known to also be a feature of *DIAPH1*-related disorder (*DIAPH1*)[9-11]. Amongst the platelet function disorders, severe bleeding and reduced platelet responses to multiple activating agonists are typically associated with Glanzmann thrombasthenia in which there is defective platelet surface expression of the αIIbβ3 integrin. However, these may also be features of several more recently discovered disorders such as CalDAG GEFI deficiency (*RASGRP2*) and leukocyte

adhesion deficiency III (*FERMT3*) in which there are defects in proteins that are functionally associated with the αIIbβ3 integrin[12,13].

Conversely, different variants in some individual genes may be associated with markedly different phenotypes of IPD. For example, in *MYH9*-RD, variants affecting the head domain of the non-muscle myosin heavy chain IIA (NMMHC-IIA) protein were found to be associated with a higher incidence of nephropathy and deafness than variants affecting the tail domain[14]. Further analysis of a larger collection of patients with *MYH9*-RD has revealed that specific variants within the same domain of NMMHC-IIA result in altered expressivity of the disorder, even when different variants affect the same amino acid residue[15].

In order to overcome these challenges, genetic testing strategies have progressed from using characteristic IPD phenotype to select individual candidate genes for confirmatory testing. Instead, new technologies now increase the scope of genetic diagnosis by enabling simultaneous examination of panels of known IPD genes or even identification of new IPD genes using more expansive sequencing approaches techniques coupled with bioinformatic and statistical genetic analyses.

DETECTION OF PATHOGENIC VARIANTS IN INHERITED PLATELET DISORDERS

Sanger sequencing

In this approach, individual genomic DNA sections of typically 100-1000 base pairs are amplified by the polymerase chain reaction (PCR) using locus-specific primers designed to flank regions of the gene of interest[16]. The PCR amplicons are then purified before chain termination sequencing to generate individual sequence traces corresponding to each PCR amplicon[16-18]. Using Sanger methodology, candidate causal variants are identified by comparing single sequence reads to reference sequence.

Although Sanger sequencing has historically been the reference standard method for detection of pathogenic variants in IPD, the technology is necessarily time consuming and has high cost. This restricts the application of Sanger sequencing to small genomic regions containing individual genes of interest. Therefore, Sanger sequencing is best suited to situations where single or small numbers of candidate genes can be selected because of characteristic clinical or laboratory phenotypes that are highly suggestive of a specific disorder. In these situations, achieving genetic diagnosis by Sanger sequencing can be seen as confirming the clinicopathological diagnosis already reached on the basis of phenotype testing alone.

High throughput sequencing

High throughput sequencing (HTS, also referred to as next generation sequencing or massively parallel sequencing) is the overarching term used to describe modern sequencing technologies using platforms such as Illumina[19], Roche 454[20] and Ion Torrent[21]. These technologies all utilise fragmented genomic DNA samples that undergo an amplification step using PCR-based

technologies. However, the main differences between the technologies are in the way that nucleotide sequence is detected and outputted. For clinical applications, it is convenient to divide the HTS approaches into gene panels and whole exome sequencing (WES) which target particular areas of the genome, and whole genome sequencing (WGS) which is not restricted to specific regions.

Gene panels

Gene panel sequencing involves the targeted sequencing a group of typically 10-100 genes that are associated with a particular disease or phenotype. In most examples, bait panels of oligonucleotide primers are designed to target the exons, flanking intronic sequence that includes the splice regions and selected regulatory regions of the relevant genes. This approach is usually applied for diagnosis of monogenic disorders for which there are several candidate genes that cannot easily be resolved using clinical and laboratory phenotype (Table 1). For example, autosomal dominant macrothrombocytopenia without additional syndromic features may be associated with variants in several genes, including *ACTN1*, *FLNA*, *GP1BB*, *TPM4* and *TUBB1.* It may be possible to distinguish some of these alternate diagnoses using phenotype tests such as flow cytometry or platelet ultrastructural analysis[4,22,23]. However, these tests are not widely available in clinical laboratories, are expensive, and are poorly standardised at present. In contrast, sequence analysis of a panel of multiple genes associated with macrothrombocytopenia is faster and more cost effective than Sanger sequencing of each individual gene.

Gene panels usually allow deeper coverage of the regions of interest for a given number of reads than can be achieved using WES, thereby increasing diagnostic accuracy. Custom bait designs for gene panels can also enable higher quality sequencing of otherwise inaccessible genomic regions that are difficult to evaluate using non-custom bait libraries such as whole exome panels. A significant economic benefit of gene panel sequencing is that it facilitates multiplexing, where several samples are sequenced simultaneously. However, one key disadvantage is that since the bait panels are custom designed to include only the genes of interest, the approach is not resilient to discoveries of new genes associated with a given phenotype. Consequently, gene panels typically require redesign to include new implicated genes, particularly in areas of frequent gene discovery.

The panel approach has been applied for genetic diagnosis of several groups of heritable disorders, including monogenic forms of dyslipidaemia[24], cardiac dysrhythmias[25] and anaemia[26]. For the IPD, the ThromboGenomics gene panel has been successfully validated in a range of different IPD and has now entered clinical diagnostic practice in the UK for analysis of IPD genes alongside other genes implicated in bleeding and thrombotic disorders[27]. Although the sensitivity (>90%) and specificity (>99.5%) of the original ThromboGenomics panel was high when considering cases with a suspected molecular aetiology, it remains to be seen whether this will be maintained as more potentially causal genes are added to the panel, and as access to the panel is broadened. The current ThromboGenomics panel (version TG3.0)

includes the coding and key regulatory regions of 100 genes associated with 58 disorders, including disorders of the vessel wall that may cause bleeding such as hereditary haemorrhagic telangiectasia and some subtypes of Ehlers Danlos syndrome (http://thrombo.cambridgednadiagnosis.org.uk/gene-disorder-list/). The content of the panel is reviewed at regular intervals and variants are reported according to American College of Medical Genetics and Genomics (ACMG) guidelines[28] by a multidisciplinary team that includes a laboratory scientist and a clinical haemostasis expert. Candidate causal variants are then validated using Sanger sequencing to ensure compliance with current UK diagnostic quality guidelines.

Elsewhere in the UK, HTS using gene panels has been applied as a research tool in the Genotyping and Phenotyping of Platelets (GAPP) study as a means of identifying new candidate genes for IPD[29]. A similar panel has been developed by researchers in the Iberian peninsula[30], and includes the exons, untranslated regions and flanking sequence of 72 genes linked to known IPDs or which are otherwise believed important in platelet biology.

Whole exome sequencing

For WES, there are several commercially available bait libraries that enable capture and amplification of the exons and short intronic flanking sequences including splice sites of all coding genes, plus other relevant non-coding functional sequence. WES bait libraries typically capture 30-60 Mb of sequence corresponding to 1-2% of the genome, although there are minor differences in the coverage between the commercial exome capture

libraries[31]. Compared to gene panels, the greatly increased number of genes that are covered by WES potentially increases the repertoire of disorders that can be detected in each analysis, primarily by enabling gene discovery in parallel with diagnosis. However, this also results in a greatly increased number of observed bystander variants that must be distinguished from potentially causal IPD variants. For WES bait libraries, this is typically in the region of 20,000 single nucleotide variants per individual, but with greater numbers observed in non-Caucasian subjects[32]. Bystander variants may include potentially pathogenic variants in targets such as cancer susceptibility genes which may be unrelated to IPD, but which may have other health impacts.

Several different research groups have utilised WES has a research tool to identify pathogenic variants in patients with undiagnosed IPD. This included a subgroup of cases recruited to the UK GAPP study[29,33,34] and the pilot phase of the NIHR BRIDGE (www.nihr.ac.uk/about-us/how-we-are-managed/our-structure/infrastructure/bioresource.htm) Bleeding and Platelet Disorders study[35]. Examples of new genes identified by WES in IPD pedigrees include *SLFN14* which results in bleeding resulting from thrombocytopenia and an additional platelet secretion defect[36]. After confirmation that *SLFN14* variants underlie IPD in other study collections, this initial WES discovery has now enabled this gene to be included as a target gene on the ThromboGenomics clinical diagnostic panel. A similar diagnostic HTS approach developed in Scandinavia utilises WES and analysis of 87 genes implicated in inherited bleeding disorders including IPD[37].

Whole genome sequencing

WGS refers to analysis of the entire genome, including coding regions and non-coding regions that include proximal and distal regulatory elements. There is emerging evidence for involvement of variants in non-coding regions in human disease[38] and advances in the annotation of the non-coding regions now makes detection of pathogenic variants more achievable[39]. WGS avoids the need for exon capture, which may be inefficient in some regions and result in coding sequence variants being missed by WES or gene panel sequencing. Although initially prohibitively expensive, the cost of WGS has now fallen to less than 1000 USD per genome[40], making WGS an increasingly cost-effective option in the diagnostic as well as research setting. Moreover, WGS utilises a single genome capture library that can be used for all disease or phenotype groups, thereby enabling better standardisation and quality assurance. Similar to WES, a single WGS analysis yields a data set for individual patients that can be stored and re-evaluated if there are new relevant gene discoveries, without the need for further analysis of samples in the laboratory.

WGS has been adopted widely in large-scale gene discovery programmes such as the NIHR BioResource - Rare Diseases (https://bioresource.nihr.ac.uk/rare-diseases/rare-diseases/) and the UK 100,000 Genomes Project (www.genomicsengland.co.uk). WGS has already enabled the discovery of several important new IPD, including platelet number disorders caused by variants in *DIAPH1*, *TPM4,* and *SRC*[10,41,42]. Other

discoveries enabled by WGS include definition of new modes of inheritance of an IPD (autosomal dominant thrombocytopenia with *GPIBB* variants) and extensions of genotype-phenotype relationship in other disorders such as CalDAG GEFI deficiency (*RASGRP2*) [43,44]. Although the key advantage of WGS at this stage is in gene discovery, as the proven mutational spectrum of disease-associated variants expands into the non-coding regions of the genome, WGS will be the only practical way to detect all potentially relevant variants.

The capacity of WGS to detect non-coding variants enables detection of causal variants underlying most forms of Thrombocytopenia absent radius syndrome (*RBM8A*) and Quebec platelet disorder (*PLAU*)[45,46], which are associated with variants in regulatory regions outside coding exons. Non-coding regulatory regions are also good candidate regions for causal variants in the approximately 40% of Mendelian IPD which cannot currently be assigned to specific genes[47]. This is particularly relevant for some sub-groups of IPD such as non-syndromic disorders of platelet secretion, for which causal genetic variants are usually elusive using current analysis approaches[7].

Detection of large structural variants and complex rearrangements

A small subgroup of IPD may be associated with large copy number variants (CNVs) or complex structural rearrangements at chromosomal level. Examples include Paris-Trousseau thrombocytopathy which may be a feature of Jacobsen syndrome (haploinsufficieny of *FLI1* because of interstitial terminal deletions of 11q)[48] or Di George's syndrome (haploinsufficincy of

ITGBA because of deletions that include 22q11.2)[49]. These diagnoses are often suspected because of characteristic syndromic features and are usually *de novo* presentations in families. However, gene panels or WES may be unsuitable because in these technologies sequence reads are not contiguous over large genomic regions.

Detection is usually achieved using a technique such as array comparative genomic hybridisation (aCGH) in which fluorophore labelled fragmented DNA samples are hybridised to an immobilised library of overlapping reference genomic DNA fragments of 100-200 kilobase pairs. In most clinical diagnostic applications of this approach, copy number variants in the range 5-10 kilobases may be detected by demonstrating failure of hybridisation with specific targets[50]. However, there is also now emerging data suggesting that WGS enables detection of large copy number variants by demonstrating changes in read coverage in either deleted or duplicated regions compared with reference sequence[51,52].

ANALYSIS OF HIGH THROUGHPUT SEQUENCE DATA

Although HTS has revolutionised diagnostic genetic testing and gene discovery in many rare disorders, resolution of causal genetic variants from irrelevant bystander variants remains a potential barrier to widespread implementation. In parallel with the development in HTS technologies, there have been other significant advances in analysis of phenotype and genotype data that have improved the utility of HTS for genetic diagnosis and for new gene discovery.

Systematic description of phenotype

In order to facilitate interrogation of HTS data, there have been several initiatives to standardise description of phenotypes. One such approach that has been successfully applied to IPD and other rare disorders is the Human Phenotype Ontology (HPO), which is a system of descriptive terms for clinical and laboratory phenotypes, but also additional characteristics such as pattern of inheritance and age of onset[53]. One important attribute of the HPO system is that it includes both general terms or leading classes (e.g. *Abnormality of blood and blood forming tissues*) and specific terms (e.g. *Absence of alpha granules*) structured in a hierarchical way so that complex phenotypes can be described in strings of terms to facilitate comparison between individuals (Figure 1) [54,55].

Annotation of IPD phenotypes using HPO terms enables streamlining of data sharing between research groups, genotype-phenotype databases, biobanks and clinical registries. HPO coding is also now used in clinical diagnostic services such as the ThromboGenomics panel, where HPO coded phenotypes assist standardisation of variant calling in the panel genes[27]. HPO terms have also been adopted by the UK 100,000 Genomes Project, as a way of helping automated selection of which panels of genes are prioritised within patients' WGS sequence datasets for reporting.

In gene discovery projects HPO terms from different individuals can be compared using statistical genetic approaches to identify groups of cases with

similar phenotypes, even though similarity may not be evident considering cases in isolation. This is exploited in statistical genetic analysis techniques such as *similarity regression* which assumes that groups of unrelated cases with HPO terms that are mathematically similar, but dissimilar to other cases, are likely to have the same underlying genetic disorder[35 56]. HPO terms in cases can also be compared mathematically to animal genetic disease models described using Mammalian Phenotype Ontology (MPO) terms to assist identification of new human disease candidates in orthologous genes. Examples of new IPD successfully identified using HPO-based similarity regression and MPO ortholog searching include *DIAPH1*-related disorder[10] and thrombocytopenia with myelofibrosis and bone defects caused by variants in *SRC*[42] respectively.

Variant prioritisation

Irrespective of the choice of sequencing technology, candidate causal genetic variants can only be identified after comparison to the reference human genome sequence at that locus. For Sanger sequencing of a single candidate gene, this is relatively simple using alignment software such as NCBI BLAST (Basic Local Alignment Search Tool)[57]. However, for HTS techniques, the complexity of sequence data requires other sequence alignment approaches and bioinformatics tools to enable calling of variants[27,30,34,35,37]. These include important considerations about sequence read quality and coverage of target regions to ensure diagnostic accuracy, particularly for clinical diagnostic services[58].

In order to help resolve candidate causal variants from bystander variants in large HTS datasets, sequence analysis requires several stages of variant filtering. This usually includes elimination of observed variants that are common or low frequency in population datasets, preferably derived from cohorts of the same ethnicity as the patient. The rationale for this approach is that the majority of pathogenic variants are rare in human populations because of negative selection pressure. Therefore, a very uncommon disease such as an IPD is most likely to be caused by a rare variant (allele frequency <0.001), particularly for dominant diseases. Commonly used resources to determine the population variant frequencies are the 1000 Genomes, UK10K and GnomAD (ExAC) datasets which contain a mixture of WES and WGS data generated through disease-specific and population genetic studies[32,59]. Care is needed to ensure that the frequency threshold for eliminating variants is appropriately set, particularly for recessive disorders in which homozygous or compound heterozygous inheritance of low frequency or even some common variants may be sufficient for disease. Amongst the IPD, this is illustrated by thrombocytopenia absent radius syndrome (*RBM8A*) in which co-inheritance of a low frequency regulatory variant and a rare variant underlies most reported cases[45].

Further semi-automated assessment of observed variants usually occurs by annotating the variants to predict the likely effect on transcript or protein with tools such as the Variant Effect Predictor. This utilises aggregated data from Ensembl and other sources to classify whether genomic DNA variants are for example, missense, synonymous coding or intronic, relative to the canonical

transcript of the gene[60]. Annotation enables variants to be prioritised for further consideration if they are predicted to have a large effect on protein expression because they are frameshift insertions or deletions, stop-gain or stop-loss missense variants, or if they disrupt splice donor or acceptor sites. Missense coding variants may also be prioritised using tools such as SIFT or PolyPhen[61,62] or the CADD score[63], which predict pathogenicity using computational criteria. Since these tools utilise different combinations of criteria to predict pathogenicity, it is usual practice to analyse candidate variants using multiple tools and progress variants for further analysis only if there is consensus between tools[37].

Gene and variant level reporting

Further shortlisting of candidate variants usually then requires consideration of several disease-specific criteria to help confirm pathogenicity. The first of these is to consider whether the variant occurs in a gene that has been previously associated with the disease phenotype of the case and with appropriate heritability (*gene level reporting*). An initiative to standardise of the selection of candidate genes for IPD and other haemostatic disorders has been initiated by the Genomics in Thrombosis and Haemostasis Scientific and Standardization Committee (SSC) of the International Society on Thrombosis and Haemostasis (ISTH) (https://www.isth.org/members/group.aspx?id=104628) and more generally in rare diseases by the and the Clinical Genome Resource (ClinGen)[64]. For example, the ClinGen proposal classifies an association between disease and a candidate gene as *Definitive* if there are numerous independent reports of

unrelated pedigrees with variants in the candidate gene in addition to strong evidence from experimental studies for causality that are upheld over time with no significant contradictory evidence. Clarification of conflicting interpretations of disease-gene associations is resolved through expert working groups[64].

Prioritised variants may then be further assessed by considering whether within implicated genes, individual variants are likely to be pathogenic (*variant level reporting*). One way of achieving this is to identify whether individual variants, or similar variants at protein level have been associated previously with the phenotype observed in the case. Publicly available variant databases are a valuable tool for variant level reporting. For IPD, these include several curated disease-specific databases such as the Glanzmann thrombasthenia and platelet type von Willebrand disease resources that are maintained under the auspices of the ISTH (https://www.isth.org/page/RegistriesDatabases). IPD data are also commercially available through more general initiatives such as Human Genome Mutation Database (HGMD)[65]. One important caution about the use of historical datasets to assist variant reporting is that some datasets may be incompletely curated and contain non-pathogenic variants that have incorrectly been associated with a disease phenotype. The potential clinical consequences of this are highly significant and include incorrect assignment of pathogenicity to observed variants in new cases. Efforts to standardise and independently curate human disease variant databases have recently been progressed by the development of the publicly accessible ClinVar variant database, which aggregates submissions from

research groups and clinical laboratories curated by panels of independent experts[66,67].

Standardisation of variant reporting

Across rare diseases, including IPD, there is an emerging priority to standardise variant level reporting to maintain quality in clinical diagnostic testing. Guidelines for the classification of variants have recently been disseminated by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMG)[28] and by the European Society of Human Genetics[68]. These emphasise that best practice for variant classification requires variant reporting in the setting of a multidisciplinary team (MDT) conference. In addition to the considerations for likely pathogenicity described above, it is recommended that additional evidence is considered including the likely functional consequences of candidate variants from experimental data and genetic considerations such as whether genotype and phenotype co-segregate within pedigrees. A scoring algorithm based on the strength of different lines of evidence then enables the simple classification of candidate variants as *pathogenic*, *likely pathogenic*, *likely benign*, *benign* or *uncertain significance*[69]. The MDT then considers whether *pathogenic* or *likely pathogenic* variants have a full contribution or partial contribution to the phenotype of the case and pedigree. One important additional recommendation is that variants of uncertain significance should not be used by clinicians for clinical decision-making. The ACMG/AMP guidelines further emphasise that variants may be reclassified and require

supplementary reporting if changes in evidence concerning the variant emerge.

FUTURE DIRECTIONS

Substantial progress has already been made in genetic diagnosis of IPD through the systematic adoption of HTS technologies for gene discovery and for clinical diagnosis pathways. Supporting these advances has been the refinement of methodologies for describing phenotype, bioinformatic analysis of sequence data and systematic evaluation of candidate variants.
Future priorities are likely to be the expansion of the diagnostic repertoire of genetic testing through ongoing gene discovery programmes and the accrual of evidence for causality and mutational spectrum of existing IPD through systematic analysis of case series. Inevitably, the focus of future gene discoveries in IPD is likely to be in non-coding genomic regions, epigenetic regulation of platelet specific genes and IPD arising as complex traits with multiple contributory pathogenic variants.

Future advances in understanding the genetic repertoire of IPD will require timely and effective translation into clinical diagnostic services for IPD, in which the emphasis will likely be delivery of cost effective and standardised delivery of laboratory sequencing services based on HTS technology underpinned by robust variant analysis and reporting processes. This approach has been pioneered by initiatives such as the UK 100,000 Genomes Project, which adopted IPD as an eligible group of disorders in 2017 for diagnosis through WGS. This initiative alongside other HTS panel services

are likely to emerge as a first line diagnostic tool for patients with suspected IPD, supported by clinical and laboratory phenotype data.

REFERENCES

1. Morison IM, Cramer Borde EM, Cheesman EJ, et al. A mutation of human cytochrome c enhances the intrinsic apoptotic pathway but causes only thrombocytopenia. Nat Genet 2008;40(4):387-389
2. Daly ME. Transcription factor defects causing platelet disorders. Blood Rev 2017;31(1):1-10
3. Pecci A, Gresele P, Klersy C, et al. Eltrombopag for the treatment of the inherited thrombocytopenia deriving from MYH9 mutations. Blood 2010;116(26):5832-5837
4. Rabbolini DJ, Chun Y, Latimer M, et al. Diagnosis and treatment of MYH9-RD in an Australasian cohort with thrombocytopenia. Platelets 2017:1-8
5. Westbury S, Downes K, Burney C, et al. Phenotype description and response to thrombopoietin- receptor agonist in DIAPH1-related disorder. Blood Adv 2018;2(18):2341-2346
6. Harrison P, Mackie I, Mumford A, et al. Guidelines for the laboratory investigation of heritable disorders of platelet function. Br J Haematol 2011;155(1):30-44
7. Hayward CP, Pai M, Liu Y, et al. Diagnostic utility of light transmission platelet aggregometry: results from a prospective study of individuals referred for bleeding disorder assessments. J Thromb Haemost 2009;7(4):676-684
8. Quiroga T, Goycoolea M, Panes O, et al. High prevalence of bleeders of unknown cause among patients with inherited mucocutaneous

bleeding. A prospective study of 280 patients and 299 controls. Haematologica 2007;92(3):357-365

9. Neuhaus C, Lang-Roth R, Zimmermann U, et al. Extension of the clinical and molecular phenotype of DIAPH1-associated autosomal dominant hearing loss (DFNA1). Clin Genet 2017; 91(6):892-901.
10. Stritt S, Nurden P, Turro E, et al. A gain-of-function variant in DIAPH1 causes dominant macrothrombocytopenia and hearing loss. Blood 2016;127(23):2903-2914
11. Ueyama T, Ninoyu Y, Nishio SY, et al. Constitutive activation of DIA1 (DIAPH1) via C-terminal truncation causes human sensorineural hearing loss. EMBO Mol Med 2016;8(11):1310-1324
12. Canault M, Ghalloussi D, Grosdidier C, et al. Human CalDAG-GEFI gene (RASGRP2) mutation affects platelet function and causes severe bleeding. J Exp Med 2014;211(7):1349-1362
13. Muller WA. Identification of a severe bleeding disorder in humans caused by a mutation in CalDAG-GEFI. J Exp Med 2014;211(7):1271
14. Pecci A, Panza E, Pujol-Moix N, et al. Position of nonmuscle myosin heavy chain IIA (NMMHC-IIA) mutations predicts the natural history of MYH9-related disease. Hum Mutat 2008;29(3):409-417
15. Pecci A, Klersy C, Gresele P, et al. MYH9-related disease: a novel prognostic model to predict the clinical evolution of the disease based on genotype-phenotype correlations. Hum Mutat 2014;35(2):236-247
16. Mullis K, Faloona F, Scharf S, et al. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol 1986;51 Pt 1:263-273

17. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74(12):5463-5467

18. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A 1977;74(2):560-564

19. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456(7218):53-59

20. Shendure J, Porreca GJ, Reppas NB, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 2005;309(5741):1728-1732

21. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011;475(7356):348-352

22. Kunishima S, Kitamura K, Yasutomi M, Kobayashi R. Diagnostic biomarker for ACTN1 macrothrombocytopenia. Blood 2015;126(22):2525-2526

23. Kunishima S, Nishimura S, Suzuki H, Imaizumi M, Saito H. TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia. Eur J Haematol 2014;92(4):276-282

24. Sadananda SN, Foo JN, Toh MT, et al. Targeted next-generation sequencing to diagnose disorders of HDL cholesterol. J Lipid Res 2015;56(10):1993-2001

25. Millat G, Chanavat V, Rousson R. Evaluation of a new high-throughput next-generation sequencing method based on a custom AmpliSeq library and ion torrent PGM sequencing for the rapid detection of genetic variations in long QT syndrome. Mol Diagn Ther 2014;18(5):533-539

26. Roy NB, Wilson EA, Henderson S, et al. A novel 33-Gene targeted resequencing panel provides accurate, clinical-grade diagnosis and improves patient management for rare inherited anaemias. British journal of haematology 2016;175(2):318-330

27. Simeoni I, Stephens JC, Hu F, et al. A high-throughput sequencing test for diagnosing inherited bleeding, thrombotic, and platelet disorders. Blood 2016;127(23):2791-2803

28. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17(5):405-424

29. Jones ML, Murden SL, Bem D, et al. Rapid genetic diagnosis of heritable platelet function disorders with next-generation sequencing: proof-of-principle with Hermansky-Pudlak syndrome. J Thromb Haemost 2012;10(2):306-309

30. Bastida JM, Lozano ML, Benito R, et al. Introducing high-throughput sequencing into mainstream genetic diagnosis practice in inherited platelet disorders. Haematologica 2018;103(1):148-162

31. Sulonen AM, Ellonen P, Almusa H, et al. Comparison of solution-based exome capture methods for next generation sequencing. Genome Biol 2011;12(9):R94

32. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. Nature 2015;526(7571):68-74

33. Johnson B, Lowe GC, Futterer J, et al. Whole exome sequencing identifies genetic variants in inherited thrombocytopenia with secondary qualitative function defects. Haematologica 2016;101(10):1170-1179

34. Leo VC, Morgan NV, Bem D, et al. Use of next-generation sequencing and candidate gene analysis to identify underlying defects in patients with inherited platelet function disorders. J Thromb Haemost 2015;13(4):643-650

35. Westbury SK, Turro E, Greene D, et al. Human phenotype ontology annotation and cluster analysis to unravel genetic defects in 707 cases with unexplained bleeding and platelet disorders. Genome Med 2015;7(1):36

36. Fletcher SJ, Johnson B, Lowe GC, et al. SLFN14 mutations underlie thrombocytopenia with excessive bleeding and platelet secretion defects. J Clin Invest 2015;125(9):3600-3605

37. Leinøe E, Zetterberg E, Kinalis S, et al. Application of whole-exome sequencing to direct the specific functional testing and diagnosis of rare inherited bleeding disorders in patients from the Öresund Region, Scandinavia. Br J Haematol 2017;179(2):308-322

38. Spielmann M, Mundlos S. Looking beyond the genes: the role of non-coding variants in human disease. Hum Mol Genet 2016;25(R2):R157-R165

39. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. Nat Rev Genet 2010;11(8):559-571

40. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcostsdata. Accessed June 2018.

41. Pleines I, Woods J, Chappaz S, et al. Mutations in tropomyosin 4 underlie a rare form of human macrothrombocytopenia. J Clin Invest 2017;127(3):814-829

42. Turro E, Greene D, Wijgaerts A, et al. A dominant gain-of-function mutation in universal tyrosine kinase SRC causes thrombocytopenia, myelofibrosis, bleeding, and bone pathologies. Sci Transl Med 2016;8(328):328ra330

43. Sivapalaratnam S, Westbury SK, Stephens JC, et al. Rare variants in GP1BB are responsible for autosomal dominant macrothrombocytopenia. Blood 2017;129(4):520-524

44. Westbury SK, Canault M, Greene D, et al. Expanded repertoire of RASGRP2 variants responsible for platelet dysfunction and severe bleeding. Blood 2017;130(8):1026-1030

45. Albers CA, Paul DS, Schulze H, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction

complex subunit RBM8A causes TAR syndrome. Nat Genet 2012;44(4):435-439, S431-432

46. Hayward CP, Liang M, Tasneem S, et al. The duplication mutation of Quebec platelet disorder dysregulates PLAU, but not C10orf55, selectively increasing production of normal PLAU transcripts by megakaryocytes but not granulocytes. PLoS One 2017;12(3):e0173991

47. Nurden AT, Freson K, Seligsohn U. Inherited platelet disorders. Haemophilia 2012;18 Suppl 4:154-160

48. Breton-Gorius J, Favier R, Guichard J, et al. A new congenital dysmegakaryopoietic thrombocytopenia (Paris-Trousseau) associated with giant platelet alpha-granules and chromosome 11 deletion at 11q23. Blood 1995;85(7):1805-1814

49. Nakagawa M, Okuno M, Okamoto N, Fujino H, Kato H. Bernard-Soulier syndrome associated with 22q11.2 microdeletion. Am J Med Genet 2001;99(4):286-288

50. Ren H, Francis W, Boys A, et al. BAC-based PCR fragment microarray: high-resolution detection of chromosomal deletion and duplication breakpoints. Hum Mutat 2005;25(5):476-482

51. Koboldt DC, Larson DE, Chen K, Ding L, Wilson RK. Massively parallel sequencing approaches for characterization of structural variation. Methods Mol Biol 2012;838:369-384

52. Hayes JL, Tzika A, Thygesen H, et al. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. Genomics 2013;102(3):174-181

53. Robinson PN, Kohler S, Bauer S, et al. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet 2008;83(5):610-615

54. Kohler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res 2014;42(Database issue):D966-974

55. Kohler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res 2017;45(D1):D865-D876

56. Greene D, BioResource N, Richardson S, Turro E. Phenotype Similarity Regression for Identifying the Genetic Determinants of Rare Diseases. Am J Hum Genet 2016;98(3):490-499

57. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215(3):403-410

58. Rehm HL, Bale SJ, Bayrak-Toydemir P, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med 2013;15(9):733-747

59. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536(7616):285-291

60. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17(1):122

61. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4(7):1073-1081

62. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7(4):248-249

63. Kircher M, Witten DM, Jain P, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46(3):310-315

64. Strande NT, Riggs ER, Buchanan AH, et al. Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. Am J Hum Genet 2017;100(6):895-906

65. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum Genet 2017;136(6):665-677

66. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 2016;44(D1):D862-868

67. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;46(D1):D1062-D1067

68. Matthijs G, Souche E, Alders M, et al. Guidelines for diagnostic next-generation sequencing. Eur J Hum Genet 2016;24(10):1515

69. Karbassi I, Maston GA, Love A, et al. A Standardized DNA Variant Scoring System for Pathogenicity Assessments in Mendelian Disorders. Hum Mutat 2016;37(1):127-134

Figure 1. Syndromic disorders can be represented using Human Phenotype Ontology terms.

The Human Phenotype Ontology (HPO) terms that might be used to describe a hypothetical patient with *MYH9*-related disorder demonstrating macrothrombocytopenia, mild bleeding, neutrophil inclusions, chronic kidney disease, cataracts and hearing impairment. HPO terms are displayed as a directed, acyclic graph generated using the HPO browser (http://human-phenotype-ontology.github.io/tools.html). The relationships demonstrated are unidirectional *is-a* relationships. Some terms overlap between leading classes, and therefore some nodes are connected by more than two edges.

Table 1. Genes implicated in inherited platelet disorders

	Genes	**MOI**
Predominantly platelet number disorders		
Amegakaryocytic thrombocytopenia with radioulnar synostosis*	*HOXA11, MECOM*	Both AR
Autosomal dominant thrombocytopenia	*ANKRD26, CYCS, ETV6, ITGA2B, ITGB3*	All AD
Autosomal dominant macrothrombocytopenia	*ACTN1, GFI1B, GP1BA, GP1BB, MYH9, SLFN14, TPM4, TUBB1*	All AD
Congenital amegakaryocytic thrombocytopenia	*MPL*	AR
Cyclic thrombocytopenia and thrombocythemia	*THPO*	AD
DIAPH1-related disorder*	*DIAPH1*	AD
Familial platelet disorder with predisposition to AML	*RUNX1*	AD
Filaminopathy with thrombocytopenia*	*FLNA*	XR
Ghosal syndrome*	*TBXAS1*	AR
MYH9-related disorders*	*MYH9*	AD
Recessive microthrombocytopenia	*FYB*	AR
Sitosterolemia with macrothrombocytopenia*	*ABCG5, ABCG8*	Both AR
Thrombocytopenia absent radius syndrome*	*RBM8A*	AR
Thrombocytopenia and inflammatory disease*	*ARPC1B*	AR
Thrombocytopenia, anemia and myelofibrosis	*MPIG6B*	AR
Thrombocytopenia with abnormalities of skin keratinisation*	*KDSR*	AR
Thrombocytopenia with myelofibrosis and bone defects*	*SRC*	AD
Wiskott-Aldrich syndrome*	*WAS*	XR
X-linked macrothrombocytopenia with dyserythropoiesis	*GATA1*	XR
Predominantly platelet function disorders		
ADP receptor defect	*P2RY12*	AR
ARC syndrome*	*VIPAS39, VPS33B*	Both AR
Autism and dense granule abnormalities*	*NBEA*	AD
Bernard-Soulier syndrome	*GP1BA, GP1BB, GP9*	All AR
Bleeding due to glycoprotein VI deficiency	*GP6*	AR
CalDAG-GEFI deficiency	*RASGRP2*	AR
Chediak Higashi syndrome*	*LYST*	AR
Deficiency of phospholipase A2, group IVA	*PLA2G4A*	AR

Glanzmann thrombasthenia	*ITGA2B, ITGB3*	AR, AR
Gray platelet syndrome	*NBEAL2*	AR
Hermansky Pudlak syndrome*	*HPS1, AP3B1, HPS3, HPS4, HPS5, HPS6, DTNBP1, BLOC1S3, AP3D1*	All AR
Leukocyte adhesion deficiency 3	*FERMT3*	AR
Paris-Trousseau thrombocytopenia and Jacobsen syndrome*	*FLI1*	AD
Platelet type von Willebrand disease	*GP1BA*	AD
Quebec platelet disorder	*PLAU*	AD
Scott syndrome	*ANO6*	AR
Stormorken syndrome*	*STIM1*	AD
Thromboxane A2 receptor defect	*TBXA2R*	AR

List of genes causally associated with inherited platelet disorders by June 2018. Although divided into disorders primarily of platelet number or platelet function, many disorders may show a combination of these features. AD – autosomal dominant. AR – autosomal recessive. MOI – mode of inheritance. * indicates those disorders with extra-hematological clinical features.

Table 2. Comparison of available sequencing techniques

	Sanger sequencing	Gene panel sequencing	Whole exome sequencing	Whole genome sequencing
Advantages	• Highly focused • Rapid, cost-effective results if correct gene sequenced	• Focused • Reduced bioinformatic load compared to WES/WGS	• Large amount of data collected • Easy to re-analyse at future date	• Greatest amount of data collected • Easy to re-analyse at future date
Disadvantages	• Time-consuming and expensive if multiple genes require analysis	• Re-versioning required to account for advances in the field • Technical aspects of selective sequence capture	• Cost (relative, at present) • Lack of non-coding sequence • Technical aspects of selective sequence capture	• Cost (relative, at present) • Number of variants / bioinformatic load
Current key uses	• Confirmation of some phenotypic diagnoses • Co-segregation analysis within families	• Routine diagnostic work-up • Co-segregation analysis within families	• At interface between diagnosis and gene discovery	• At interface between diagnosis and gene discovery

Key advantages and disadvantages of the different sequencing techniques currently used in IPD. Selection of sequencing technique may also be influenced by local healthcare funding or commissioning strategies.